

RESEARCH

Open Access



Surfactant representation using COSMO screened charge density for adsorption isotherm prediction using Physics-Informed Neural Network (PINN)

Achmad Anggawirya Alimin¹, Kattariya Srasamran¹, Wanutchaya Yuenyong¹, Ampira Charoensaeng^{1,2}, Bor-Jier Shiau³ and Uthaiporn Suriyapraphadilok^{1,2*}

Abstract

Predicting surfactant adsorption using the currently available isotherm model is limited to one or two independent variables: equilibrium concentration and temperature. This study aims to develop an adsorption model that includes molecular features, testing conditions, and solid properties in the model. A Physics-Informed Neural Network (PINN) was structured by integrating adsorption isotherm into artificial neural networks (ANN). The model was trained using a dataset containing 56 adsorption isotherms and 20 types of anionic and nonionic surfactants under various conditions with sand and silica oxide as their solids. The surfactants were quantified using sets of descriptors generated from molecular counting, charge distribution, and Conductor-like Screening Model (COSMO) screened charge density. The COSMO-screened charge density descriptors provide the highest accuracy in representing the surfactant molecule. The interpretation of molecular structure effect and surfactant-solid interaction described using COSMO-screened charge density showed that adsorption between the surfactant and solid media involves hydrogen bonding and hydrophobic interaction. The PINN model achieves high accuracy with 93% training and 85% validation with five-fold cross-validation. Later, the model was evaluated and used to generate an adsorption isotherm and predict unseen surfactant adsorption. Adsorption prediction with unseen surfactants showed high accuracy with the surfactant for familiar structure (RMSE 0.07 mg/g) and promising profile for the whole new structure (RMSE 2.95 mg/g).

Scientific contribution This study advances the field by integrating COSMO-screened charge density descriptors into a physics-informed deep learning model to predict surfactant adsorption isotherms, accounting for molecular features, testing conditions, and solid properties. The incorporation of COSMO-screened charge density offers a novel approach to accurately represent surfactant molecules, enabling accurate prediction of their adsorption behavior. This approach extends conventional models, which are often limited to empirical parameters or fewer variables. This physics-informed framework significantly enhances the understanding of surfactant-solid interactions and offers a robust predictive tool for optimizing surfactant formulations, aiming to minimize adsorption losses in chemical enhanced oil recovery and environmental remediation.

Keywords Physics-informed neural network, Surfactant, Adsorption, COSMO, Chemical enhanced oil recovery

*Correspondence:

Uthaiporn Suriyapraphadilok
uthaiporn.s@chula.ac.th

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

A surfactant is an amphiphilic molecule with different polarities, a polar head, and a non-polar hydrocarbon tail in one structure. Due to its distinctive polarity profile, surfactant molecules are concentrated at the interface between two substances with different polarities. This unique characteristic of surfactants reduces the interfacial tension (IFT) and increases the miscibility between the two fluids, either liquid–gas or liquid–liquid phases [1]. This property has also made surfactants critical in, e.g., cleaning, environmental remediation, and enhanced oil recovery (EOR) applications, as both low interfacial tension and high solubility are required to achieve high oil recovery [2, 3]. However, the use of surfactants in reservoir rock formations or the subsurface porous media comes with a threat of material losses due to excessive surfactant adsorption.

Adsorption of surfactant onto a solid surface can be somewhat expensive and becomes a prohibitive factor of field implementation, considering the chemical cost and the surfactant quantity needed to overcome the limit of adsorption [4, 5]. There were numerous studies investigating the controlling factors of surfactant adsorption onto solid surfaces, for example, including (1) surfactant types/structures and ranges of concentration, (2) elevated reservoir temperature, (3) salinity, and (4) solution pHs [1, 6]. Among those dominant parameters, in particular, influences of concentration, temperature, salinity, and pH have been mostly emphasized in most studies previously [6–9]; generally, they can be measured experimentally and quantified accurately. However, any detailed surfactant structure effects associated with adsorption were much of a challenge to be addressed quantitatively and thoroughly.

In general, the charged surfactant head group mainly affects affinity to the solid surface; for example, Mas-sarweh and Abushaikh [2] and Kalam, Abu-Khamsin [10] have documented a thorough overview, where (i) cationic (positive charge on the headgroup) surfactant is attracted to the negatively charged rock surface, such as common sandstones, (ii) anionic (negative charge on the headgroup) surfactant possesses strong affinity toward the positively charged clay edges if present on the media, and (iii) hydrogen bonding found in ethyl oxide or propyl oxide based nonionic (no charge on headgroup) surfactants are preferably attracted to the hydrogen bonding donors present on the solid surface (i.e., sandstones). Prior study by Somasundaran, Middleton, and Viswanathan [11] has qualitatively described the structure's effect on surfactant adsorption. Using a quantitative approach, Atkin, Craig [12] offered some insights linking the adsorption data to the kinetic and structural information of the surfactant using tools such as atomic force

microscopy, fluorescence quenching, and neutron reflectivity. In brief, the surfactant head group and its structure play critical roles in affecting adsorption onto the solid surface. However, based on our knowledge, a few models use such information to predict surfactant adsorption on sand or clay media for chemical enhanced oil recovery (cEOR) or soil remediation applications.

There have been various ways to quantify the molecules and their particular behavior. One intuitive approach to better quantify the targeted molecules, widely investigated in multiple fields, is using the group contribution (GC) concept for pure components [13, 14]. The GC method estimates a property by counting the features of a molecule from the smallest section (first-order) to the functional groups (second-order), which are to be regressed in a linear model [15]. Others have studied some specific surfactant properties, and predictions based on the GC method, for example, critical micelle concentration (CMC) [16], hydrophilic-lipophilic balance [17], relative solubility number [18], cloud point [19], and Krafft point [20]. However, the GC method requires a significantly large dataset for higher accuracy. Another alternative to the GC method in quantifying the molecule is to use the charge distribution to overcome the situation of the limited dataset available. The surfactant charge distribution between the head and tail sections influences the self-assembly process [21]. This self-assembly process eventually affects the profile of the adsorption isotherm. In a more recent approach, Klamt [22] introduced an improved approach to molecular characterization using a so-called solvation model in a perfect conductor: the Conductor-like Screening Model (COSMO). In recent decades, molecule representation applying this COSMO with machine learning has become increasingly popular with different applications, e.g., prediction of properties [23], solvent selection [24], and activity coefficient [25, 26]. Although COSMO accuracy is lower in properties prediction applications than matured GC methods like UNIFAC and UNIQUAC, the modified COSMO method has continued to evolve. The latest improvement in COSMO involves different types of hydrogen bonding and dispersive interaction, COSMO-SAC [27, 28]. Also, the COSMO model has shown encouraging results in the adsorption prediction of ionic substances in various solid adsorbents, e.g., dyes, pyridinium ionic liquid, and ibuprofen with activated carbon [29–31], dyes with silica [32], dyes with organic solid [33], and database of substances with soil [34].

Several efforts have been made to predict surfactant adsorption in the adsorption application using machine learning. Surfactant adsorption onto solid surfaces can be predicted using adsorption isotherms, which provide a curve describing the number of surfactant molecules

in the aqueous phase that adhere to the solid surface at different equilibrium concentrations [35, 36]. Adsorption isotherms have been developed in various models [37] and are unique for every system. Kalam, Abu-Khamsin [38] trained machine learning models, XGBoost and random forest, to predict the cationic Gemini surfactant adsorption in chemical EOR system with the input data of mineral composition, maximum adsorption capacity, and surfactant concentration. Another approach by Chen, and Huang [39] utilized the ANN model to estimate the capacity of biosurfactant adsorption on activated carbon. The ANN approach showed promising results with more than 90% accuracy. However, both models are limited to only a single surfactant used in their study. Although the ANN model could perform well, there is no further interpretation of how the model could represent the adsorption behavior of the surfactant on a substrate. In EOR and contaminant remediation using surfactants, the reservoir and aquifer conditions vary widely in temperature, pH, salinity, and subsurface media composition. The existing isotherm model and approaches thus are severely insufficient to accommodate this wide range of variation as a single model, making the whole process time-consuming and inefficient due to countless models to manage. Thus, a model that can incorporate more independent variables simultaneously should significantly reduce the number of variations in lab testing, reducing the time and resources spent. To our knowledge, this is the first time that the PINN technique coupling the structural features from COSMO is used to predict the adsorption of surfactant with several distinct structures onto different rocks and sands media in EOR or environmental remediation applications.

This study focused on integrating dual benefits of the learning performance of the ANN model and the quantum chemical calculation based on the COSMO in the description and prediction of surfactant adsorption onto silica surface; hence, the relationships between the model developed and adsorption phenomena are concise and explainable through representative chemical interactions and forces including hydrogen bonding donor affinity, hydrogen bonding (HB) acceptor affinity, and hydrophobicity. A classical Langmuir equation will be integrated into the ANN architecture as a Physics-Informed Neural Network (PINN) model for simplification. The surfactant's molecular features were decrypted effectively using several tools of feature counting, charge distribution, and COSMO-screened charge density. The updated data was then combined and expanded with specific testing conditions (temperature, pH, salinity) and solid properties (size, specific area, and point of zero charges) to be categorized as a sub-dataset. The modified dataset is then fed into the machine learning and ranked utilizing

recursive feature elimination (RFE), a feature selection technique that iteratively removes the least important variables based on model performance. By reducing dimensionality and eliminating irrelevant or redundant inputs, RFE helps improve model generalizability and reduces the risk of overfitting. The result of the highest accuracy dataset of the PINN model was used to develop an adsorption model. The result was then analyzed, tested, and validated using unseen surfactant adsorption, as shown in the overall workflow as depicted in Fig. 1.

Methodology and approach

Dataset collection and assumptions

The dataset was harvested from various sources [7, 9, 40–46] and then categorized into parameters of experimental results, testing conditions, solid properties, and surfactant descriptors. The collected data comprised 552 equilibrium adsorption data points with 20 surfactants on sand and silica oxide surfaces. The surfactant types vary from anionic with linear chain (e.g., Sodium Dodecyl Sulfate, SDS), anionic with double branch chains (e.g., bis(2-ethylhexyl) sulfosuccinate, DHSS or AOT, and Internal Olefin Sulfonate, IOS), anionic with double headgroups (e.g., Alkydiphenyloxide Disulfonate, DOW-FAX), anionic with phenyl group(s) (e.g., Alkyl Benzene Sulfonate, ABS; Sodium Dodecyl Benzene Sulfonate, SDBS), and anionic with polyethylene oxide group (e.g., Sodium Lauryl Ethyl Sulfate, SLES). Nonylphenol polyethoxylated-based nonionic surfactants (NP) vary from alkyl chains (e.g., Alkylphenol Ethoxylate, Tergitol), and polyoxyethylene oxide head groups (e.g., Triton-X), and alcohol ethoxylated based nonionic surfactants (e.g., Alkyl Alcohol Ethoxylate or Neodol). During the model generation, 80% of the data was used for training, while 20% was used for testing through fivefold cross-validation. Four surfactant isotherms are also used to challenge the model with unseen surfactant molecules.

There are three selected parameters in the testing condition: temperature ($^{\circ}\text{C}$), salinity (mg/g), and pH of the solution. Some references did not measure all testing conditions; therefore, these data were assumed to be neutral (i.e., salinity at 0, pH of 7, and temperature at 25°C). Meanwhile, the solid properties are represented by PZC, grain size (micrometer), and surface area ($\frac{\text{m}^2}{\text{g}}$). Similar to the testing condition, some external data sets did not measure the detailed characteristics of solid media. Instead, the authors mentioned the types of solids being used. Therefore, the media properties were estimated using readily available data based on solid types. Meanwhile, the surfactant quantification is discussed in the following subsections. The solid properties, surfactant

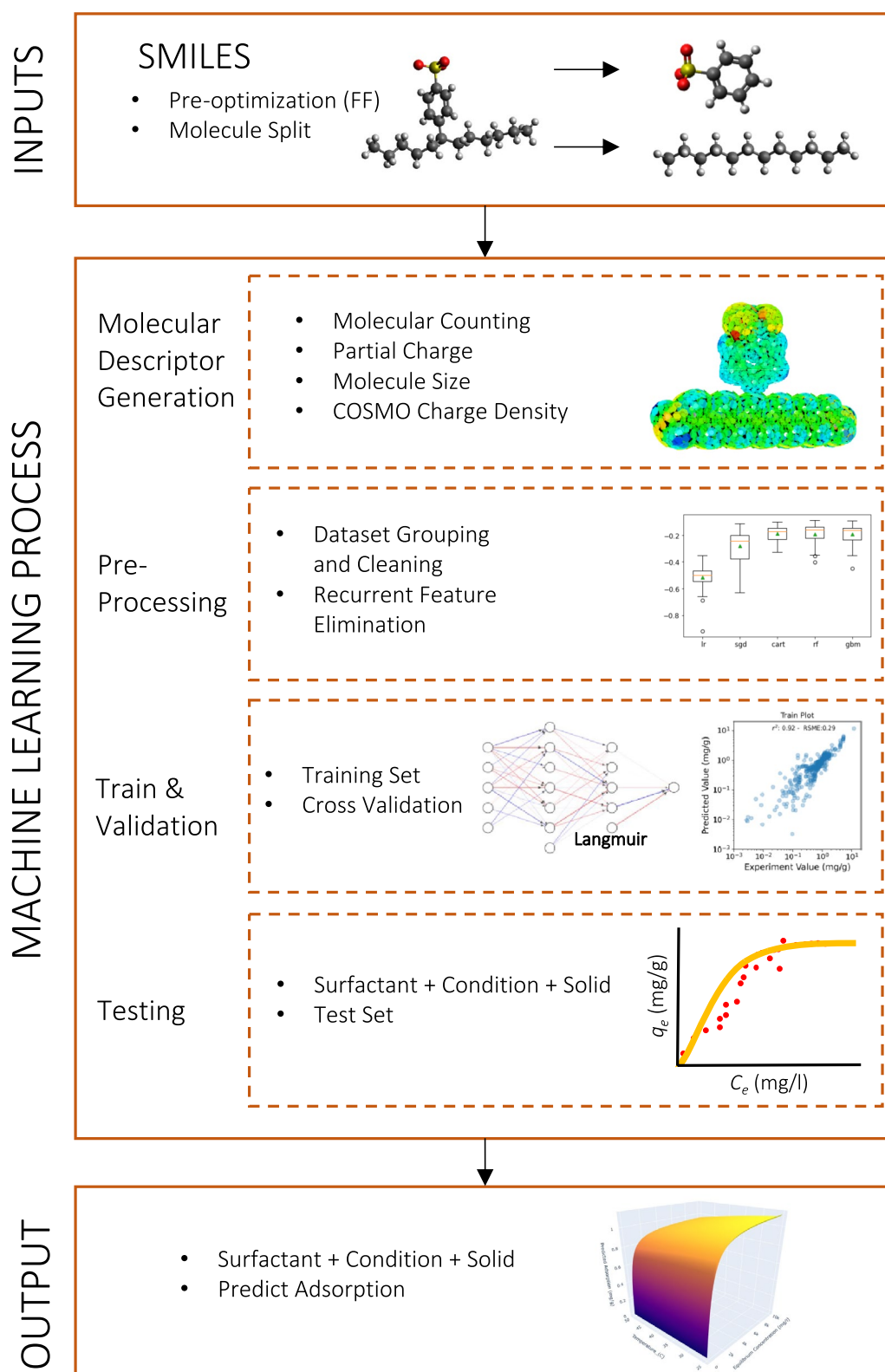


Fig. 1 Workflow schematic of descriptor generation, training, and output model

quantification, collected adsorption data, and solution conditions are provided in Supplementary Materials Table S1, S2, and Figure S2.

Surfactant quantification method

Molecular counting

Molecular counting was performed by identifying and polling all the molecule's functional groups. This study represents the functional groups by Ethylene Oxide (EO), Propylene Oxide (PO), and phenyl groups. The EO and PO groups affect the solubilization of oil, water, and surfactant as they introduce the intermediate polarity to the surfactant molecule. The solubilization increases with the higher number of extended functions [47]. For instance, the EO group increases the solubilization in the water phase, whereas the higher number of PO groups gives higher solubility in the oil phase. The study of Postmus, Leermakers, and Stuart [48] has shown that alkyl-ethylene oxide (EO) affects the aggregation of nonionic surfactants at the silica-water interface, further impacting the adsorption. Arpornpong, Charoensaeng [49] examined the adsorption characteristics of extended surfactant adsorption on alumina. The result showed that the surfactant with more EO group besides the head group resulted in higher adsorption area per molecule ($\text{m}^2/\text{molecule}$), lowering actual adsorption. A similar observation was reported by Rattanaudom, Alimin [50].

The surfactant geometry descriptors are represented by surfactant length, tail length, and types of branching. The surfactant length was calculated directly from the most extended atom sequence between both ends of the molecule. At the same time, the surfactant tail length is considered the longest carbon chain in the surfactant molecule, not including EO, PO, and benzene rings. Somasundaran, Middleton, and Viswanathan [11] found that increasing the chain length of alkyl groups increased adsorption in four regions of the adsorption isotherm except the super-micellar region [51]. Meanwhile, the types of branching were categorized into four types, from linear to highly branched molecules based on Somasundaran, Middleton, and Viswanathan [11]. Both surfactant length, tail length, and types of branching directly affect the molecular packing parameter, which defines the shape of micelles and equilibrium aggregates of the surfactant [52]. The phenyl group affects the surfactant aggregation via the π - π interaction between the benzene ring of the first layer adsorbed surfactant and surfactant molecule [53–55]. COSMO sigma profiles of the benzene ring and silica oxide generated by Man, Abdullah [56] showed the presence of both polar and non-polar charge density regions. This potentially leads to the hydrophobic and polar-polar interactions, which increase the adsorption affinity of the benzene ring onto the silica oxide.

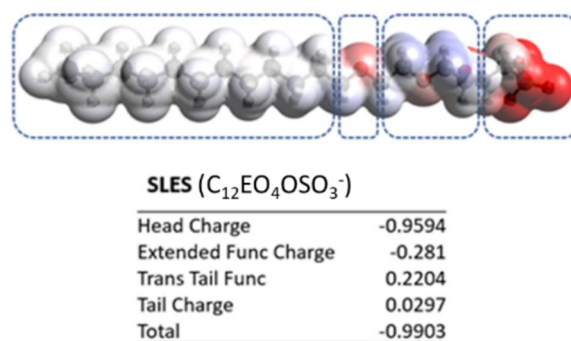


Fig. 2 Sodium Lauryl Ethyl Sulfate (SLES) molecule split into the head (OSO_3^-), extended function (4EO), transitional (C), and tail of C12; the partial charge from each atom in the respected section was accumulated as the section charge; the total charge of the whole molecule is still close to the theoretical value of an anionic surfactant (-1)

Charge distribution

The charge distribution was calculated by sectioning the surfactant molecule and adding the partial charge from each atom for each section. Huibers [21] showed that the partial charge distribution from the surfactant section influences surfactant self-assembly and physical properties. It is also suggested that the head group charge influences the molecular interaction. In this study, the surfactant molecule is separated into four sections: head, extended function, transition, and tail. The partial charge for each atom was calculated using the Gasteiger-Marsili equation [57] readily provided in the Avogadro[®] [58]. Figure 2 shows the charge segmentation of Sodium Lauryl Ethyl Sulfate (SLES).

COSMO-screened charge density

This study calculated the COSMO-screened charge density in GAMESS 2021 R2 [59] using a pre-optimized molecule from Avogadro [60]. A simplified molecular-input line-entry system (SMILES) was used in Avogadro[®] to generate the molecule. The generated molecule was then pre-optimized using the MMFF94 force field (500 steps, steepest descent, and 10^{-7} as convergence). The output from this section was pre-optimized to relatively locate each atom in a discrete vacuum space (x, y, z-axis).

In GAMESS, molecule optimization configurations were calculated with the atom's relative location. The selected wave function was Restricted Hartree-Fock (RHF), with the density functional theory of BP86 and KTZVP as the Gaussian basis. Meanwhile, the perfect dielectric constant ($\epsilon = 80$) was used to simulate a water-equivalent virtual continuum. The internal charge of molecule "icharge" was set as -1 for anionic and 0 for nonionic surfactant. GAMESS calculation generated

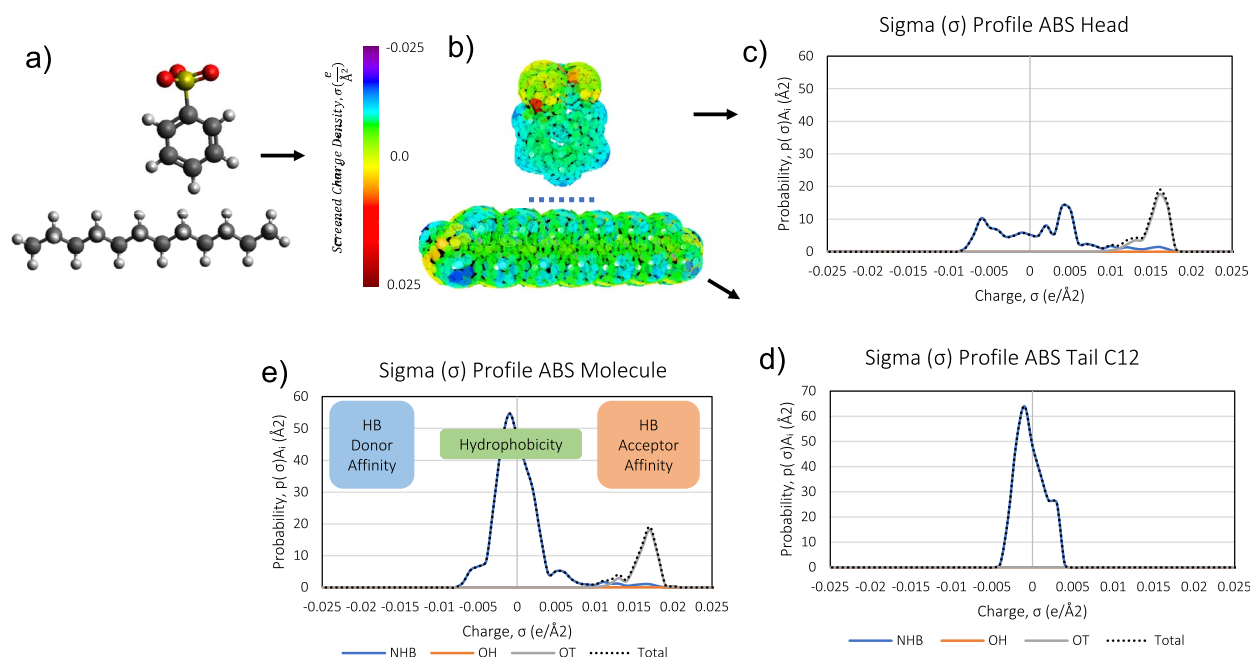


Fig. 3 The COSMO molecular descriptor steps from generating and splitting the alkyl benzene sulfonate (ABS) molecule from Avogadro[®], generating the screened charge density in GAMESS, and plotting the sigma profile. The sigma profile shows the probability of finding a specific charge on the molecule surface. Peaks from two regions (hydrophobicity and HB-acceptor affinity) are identified. At the same time, the charge comes from OT, OH, and non-hydrogen bonding (NHB) atoms

hundreds to thousands of charged density sections wrapping the molecule. An open-sourced Python code was deployed to consolidate the charge density into the sigma profile by normalizing and plotting the charge density with its probability. The GAMESS configurations and Python codes for sigma profile generation were sourced from Bell, Mickoleit [61].

The generated sigma profiles, as shown in Fig. 3, present the charge distribution and probability of finding specific charge values. The charge value in the sigma profile was classified into three regions: hydrogen bonding donor (HB-donor) affinity, hydrogen bonding acceptor (HB-acceptor) affinity, and hydrophobicity [62]. The HB-donor affinity, which ranges the sigma potential (σ) from -0.025 to -0.01 e/Å², shows the tendency of a specific molecule surface to patron its hydrogen. Meanwhile, the HB-acceptor affinity with σ from 0.01 to 0.025 e/Å² shows the tendency of a specific molecule surface to accept hydrogen. HB-acceptors generally come from electron valence in O, N, and F atoms, which are further classified into OH if the valence comes from an oxygen-hydrogen (hydroxyl) bond such as OH function in carboxylated surfactant, and OT (other) if the valence comes from O, N, and F with other atoms. For example, in alkyl benzene sulfonate (ABS), the acceptor region is provided by the oxygen atom (OT) in the sulfonate headgroup. Furthermore, the hydrophobicity region, which ranges the sigma potential (σ) from -0.01 to 0.01 e/

Å², reflects the neutral surface area. This segment typically shows the hydrocarbon tail(s) of the molecule or neutral segment between the HB-donor and HB-acceptor.

The sigma profile was consolidated into a discrete value following the approach proposed by Lukowicz, and Benazzouz [23] to calculate charge segmentation from each HB-donor, HB-acceptor, and hydrophobicity region. Different bonding from different atoms results in different interaction strengths, so the COSMO-SAC models [27] differentiate the interaction based on the contributing atom (OH, OT, and NHB). In this study, the sigma profile was integrated based on charge segmentation and atom contribution. Therefore, nine different values of molecular descriptors were generated from a single sigma profile. Equation 1 shows the sigma profile integration of OT for each density region.

$$OT_{acceptor} = \int_{0.01}^{+\infty} p^X(\sigma_{OT}) d\sigma \quad (1)$$

$$OT_{hydrophobicity} = \int_{-0.01}^{0.01} p^X(\sigma_{OT}) d\sigma \quad (2)$$

$$OT_{donor} = \int_{-\infty}^{-0.01} p^X(\sigma_{OT}) d\sigma \quad (3)$$

Equations 1, 2 and 3 show the integration of the sigma profile of three different sections from the interaction of OT atom, where $OT_{acceptor}$ is the cumulative charge of hydrogen bonding acceptor (HB-acceptor) region from OT atoms in e , $OT_{hydrophobicity}$ is the cumulative charge of hydrophobicity region from OT atoms in e , OT_{donor} is the cumulative charge of hydrogen bonding donor region (HB-donor) from OT atoms in e , σ is the screened charge density for each section in $\frac{e}{\text{\AA}^2}$, and $p^X(\sigma_{OT})$ is the segment area covered by screened charge density σ in $\text{\AA}^2$.

Since surfactant adsorption is highly affected by head and tail orientation during adsorption, the head and tail sections were simulated separately to reveal the adsorption of each segment in addition to the overall surfactant molecule. The molecule split is also by considering that the COSMO cavity in GAMESS is limited to 150–200 atoms. Therefore, three different sigma profiles were generated for each surfactant molecule: (1) the head section with extended function, (2) the tail section, and (3) the whole molecule. The steps from the input data, molecule split, calculation of screened charge density, generation of sigma profile, and the integration are shown in Fig. 3.

Descriptors selection and model training

Recursive feature elimination (RFE) was employed to evaluate the performance of each given dataset. The RFE used in this study was taken from the Python Scikit-Learn [63]. The tools take several descriptors simultaneously, train them in the machine learning algorithm, and compare them to another result set with slight changes by changing the descriptors. This algorithm introduces the generalization in a problem with a small number of datasets. The generalization will also vanish weak features, which did not significantly improve the model accuracy [64].

The RFE study conveniently splits datasets into five groups based on distinct surfactant quantifying methods. Each dataset must contain at least testing conditions (temperature, pH, and salinity), solid properties (PZC, specific area, and particle size), and different sets of surfactant descriptors. The first group used all generated descriptors (molecular counting, charge distribution, and COSMO) as surfactant descriptors. The second group used molecular counting only, and the third used charge distribution only. The fourth group used the whole molecular COSMO descriptors, and the fifth used the split (head and tail) portion of COSMO descriptors. The highest performance set of descriptors is then used to train the PINN model. The selection using RFE is described further in Supplementary Material. By selecting only the most informative features, RFE can reduce the sensitivity of the model to small fluctuations in the

training data, simplify the model, and mitigate the risk of overfitting.

A PINN model was generated using a functional application programming interface (API) and sub-classing API provided by Tensorflow [65] and Keras [66] libraries in Python. The functional API enabled the passing of the equilibrium concentration to the last layer. The sub-class API was utilized to develop a custom neuron with the Langmuir equation introduced at the last layer in the neural network architecture. Incorporating a physics-based model into the backpropagation process constrains the neural network to adhere to underlying physical laws. This acts as an implicit regularizer, guiding the learning process and helping to reduce overfitting. The generated model was then trained with a previously selected dataset from RFE, with 80% training and 20% for validation with fivefold cross-validation. Figure 4 shows the neural network configuration for the PINN. The RMSE was chosen as the loss function, as the training algorithm is expected to predict higher concentrations better. This selection is due to the nature of adsorption behavior, as the most important adsorption values are located at their highest saturation region where surfactant molecules are fully adsorbed on the solid surface. The RMSE loss function will punish the inaccuracy of outliers more severely, which leads to the algorithm suitable for the adsorption isotherms that give good accuracy at maximum adsorption values. Further, the converged model was evaluated and tested using unseen surfactant adsorption data. Finally, SHAP (Shapley Additive explanations) analysis is deployed to evaluate the impact of each descriptor on the model.

Results and discussion

Evaluation of the PINN model

The RMSE was chosen as the loss function, as the training algorithm is expected to better predict higher concentrations. This selection is due to the nature of adsorption behavior, as the most important adsorption values are located at their highest saturation region where surfactant molecules are fully adsorbed on the solid surface. The RMSE loss function will punish the inaccuracy of outliers more severely, which leads to the algorithm suitable for the adsorption isotherms that give good accuracy at maximum adsorption values. As shown in Table 1, the average R^2 of the training dataset was 93% and 84% for validation, while the RMSE for training was 0.28 and 0.40 for validation. Therefore, each fold of the five-fold cross-validation result indicates that the model is consistent throughout the dataset. Fold number Two has shown the highest R^2 training and testing score out of the five models. This fold has a better data distribution with noise in

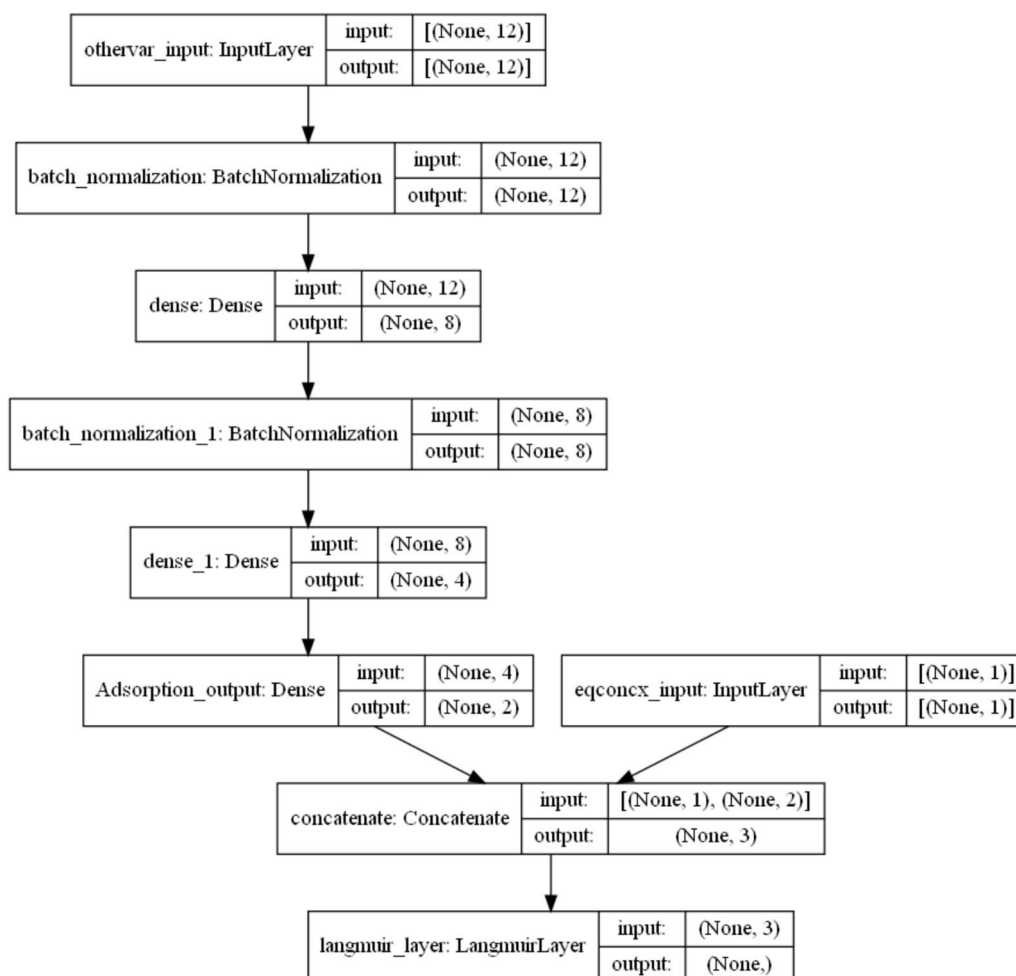


Fig. 4 Sub-Classing Langmuir equation into Neural Networks

the training and validation set than other folds; therefore, the accuracy between training and testing is close.

The PINN model accuracy in the dataset was evaluated using scatter plots to visualize the error distribution between experimental and predicted values; the result is shown in Fig. 5. The ideal prediction would align all the dots in the perfect diagonal line. The resulting plot shows a tapered trend toward higher adsorption since RMSE was selected as a loss function. Although the accuracy (%) showed a high result, the absolute value of adsorption error at higher concentrations tends to be high.

Figure 5 shows that the outliers are mostly sulfate surfactant (red, RMSE 1.39 mg/g), followed by sulfonate (blue, RMSE 0.39 mg/g) and alcohol ethoxylate (green, RMSE 0.26). The high errors could originate from the purity and heterogeneity of the surfactant used in the experiment. The purity of as-received surfactant could range from 25 to 30% up to 99% active components. Most surfactants have been produced from specific

Table 1 Five-fold cross-validation results PINN model using COSMO head and tail descriptors; the visualization for error distribution in each fold is provided in Supplementary Material Figure S4

Fold number	R ² training (%)	R ² validation (%)	RMSE training	RMSE validation
0	94.1	74.2	0.29	0.37
1	91.6	84.7	0.28	0.59
2	94.3	91.4	0.24	0.43
3	92.8	83.6	0.32	0.30
4	94.2	85.8	0.29	0.30
Average	93.4	83.9	0.29	0.40

ranges of alkyl chain length, such as the tail part, e.g., C12-C14 and C14-C18, depending on the hydrocarbon ranges. The heterogeneity of ethylene oxide or propylene

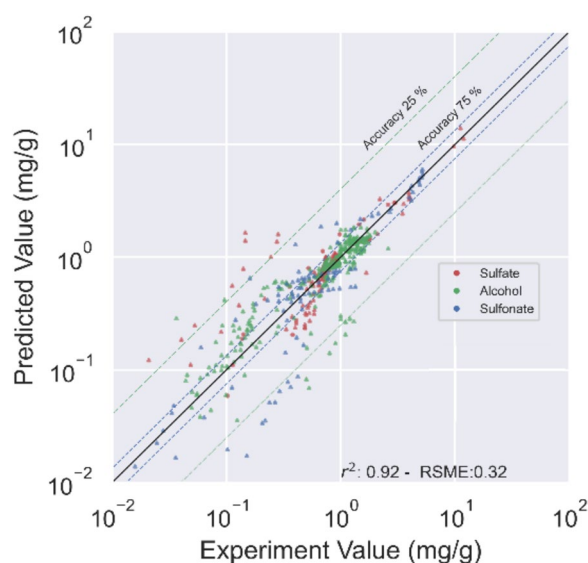


Fig. 5 Parity plot between the predicted surfactant adsorption capacity obtained from the PINN model vs. the experiment values with various surfactant headgroups onto the sand surface

oxide groups in the ethoxylated nonionic surfactants or different positions of the branching alkyl groups also contributes to surfactant polydispersibility. Besides surfactant and solution characteristics, solid morphological and mineralogical characteristics also affect surfactant adsorption [4]. Sandstone is an agglomeration of individual minerals consisting of a primary component such as quartz and other minerals, including chert, feldspar, mica, illite, kaolinite, and calcium carbonate [67]. Depending on the degree of heterogeneity, all these factors have contributed to the deviations from the model prediction.

Physical meaning of selected descriptors

Recursive feature elimination (RFE) was employed to evaluate the performance of each given dataset, and the detailed analysis is described in Supplementary Material Figure S1. The RFE narrowed the descriptor choices from 26 COSMO head and tail descriptors set into 13. We have also examined that as the number of features tested increases, the model's accuracy increases until the specific number of descriptors. The 13 selected features represent variables that matter to the phenomena. However, the model accuracy eventually reached a plateau as an additional feature did not contribute to higher accuracy. The plateau indicates that the additional features did not correlate to the predicted values or the feature already represented by another feature already included in the descriptor set. Out of the 13 features selected, one represents equilibrium concentration, three represent

the testing condition (temperature, salinity, and pH), two represent the solid characteristics (e.g., PZC and particle size), and the other seven descriptors represent the molecular descriptors. The molecular descriptors are as follows: five COSMO descriptors (i) NHB-HB-donor, (ii) Head section NHB-Hydrophobicity, (iii) Head section NHB-HB-acceptor, (iv) Head section OT-HB-acceptor, (v) Tail section NHB-Hydrophobicity, and two general molecular properties, (vi) Molecular volume, (vii) Molecular length.

Based on selected COSMO descriptors, the surfactant descriptors cover all types of bonding affinity from HB-donor, HB-acceptor, and hydrophobicity. The hydrogen bonding donor and acceptor appear from the head section of the molecule where the oxygen atom with valence is found. This is confirmed by selected COSMO descriptors i, iii, and iv. Meanwhile, the hydrophobicity is represented by the ii and v COSMO descriptors, where the Head section NHB-Hydrophobicity represents the hydrophobicity from benzene or extended function in the headgroup, and the Tail section NHB-Hydrophobicity represents hydrophobicity from carbon chain in the tail section. Tail NHB-hydrophobicity represents the surfactant's neutral charge area on the tail section. NHB stands for non-hydrogen bonding, which comes from alkyl-alkyl interactions. The longer the tail, the larger its NHB-hydrophobicity value is observed.

On the other hand, the atomic interaction only covers the NHB and OT-types of hydrogen bonding, which represent hydrogen bonding affinity from other atoms besides hydroxyl. OT means it comes from other hydrogen bonding, such as oxygen from sulfonate, sulfate, or alcohol (other than hydroxyl).

Interpreted interaction between surfactant and silica surface: a case study of SDBS and SiO_2

The possible interaction was investigated by examining the screened charge density for possible interactions between surfactant and silica surface. The surfactant selected to be compared was SDBS, while the surface was silica oxide (SiO_2), generated from SiO_2 lattice crystals (two by two by one). The number of hydroxyl functions was then adjusted to 6, referring to the specific number of hydroxyl per surface area at maximum hydration [68]. Maximum hydration was chosen following the reservoir condition during chemical flooding, where a mixture of water and surfactant was used to sweep the reservoir. The surface charge of the SiO_2 molecule was then adjusted with a formal charge of -3 , expressing the negative surface charge following the reported experimental PZC value of silica and quartz between 2 and 4 [69, 70]. Similar to the surfactant, the molecule was generated

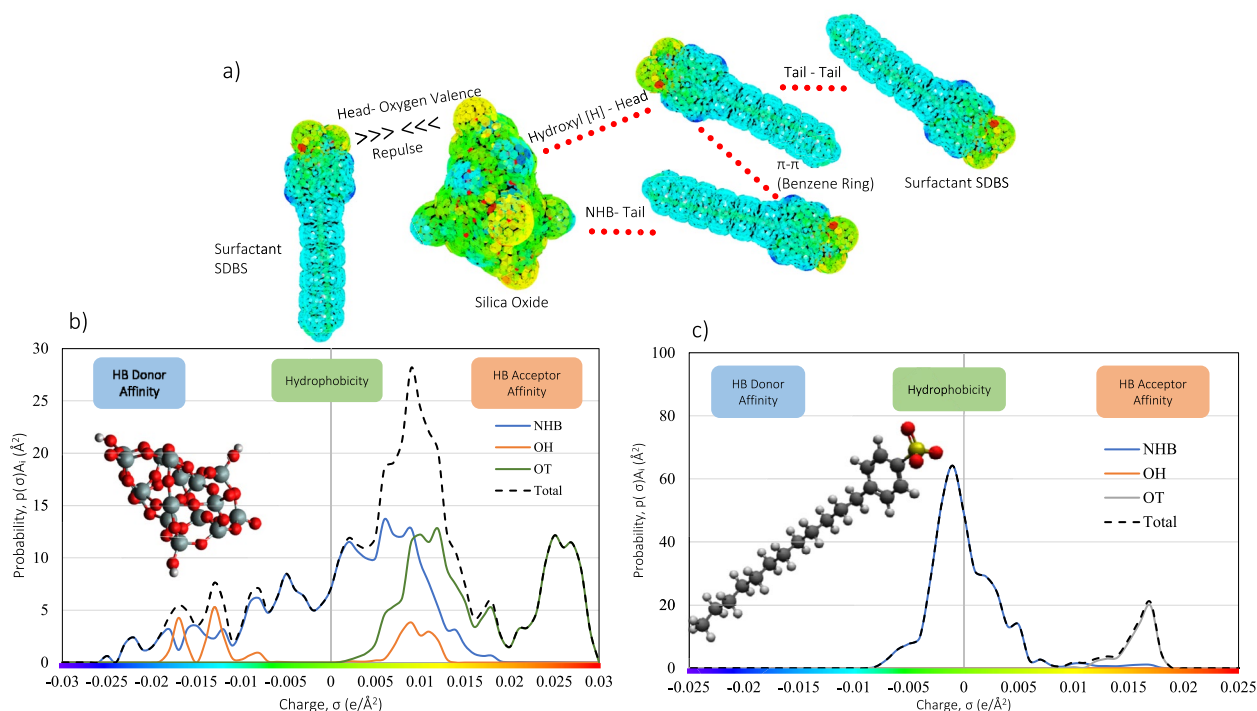


Fig. 6 **a** COSMO-screened charge density and probable interaction of surfactant Sodium Dodecyl Benzene Sulfonate (SDBS) and silica oxide, **b** sigma profile silica oxide, and **c** sigma profile surfactant SDBS

in the Avogadro[®] using lattice builder and exported to GAMESS for screened charged density calculation.

The COSMO-screened charge density results and interaction interpretations are shown in Fig. 6a. The figure shows the surfactant and SiO₂ molecule wrapped by thousands of tiny pieces of COSMO-screened charge density. Each section is color-coded based on the charge value, ranging from blue for negative charge, green for neutral charge, and red for positive charge. It should be noted that the screened charge shown is the charge from a continuum perspective. This implies that the negative values on the molecule surface are expressed as a positive charge in the continuum cavity.

As illustrated in Fig. 6b, SiO₂'s wide variety of functional groups enable the σ -profile to span into all three types of charge region: (1) non-hydrogen bonding, which is caused by neutral SiO₂ atoms; (2) HB-donor, which comes from hydrogen in the hydroxyl (silanol) function; and (3) HB-acceptor, which comes from two pairs of valence electron at the oxygen atom in SiO₂. According to this, all hydrophobic and hydrogen bonding interactions are theoretically feasible on the surface of SiO₂, with the strength of each contact being determined by the quantity of hydroxyl functional groups, in this case, the degree of hydration. The HB-acceptor peaks at about 0.01 $e/\text{\AA}^2$ towards the neutral charge area, which makes

the molecule more negative, illustrating the impact of the bulge hydroxyl function, making the whole molecule more polar. Das and Bhattacharyya [71] confirmed that SiO₂ transfers hydrogen bonding potential from silanols to give bonding potential. Through donor–acceptor interactions, the polar hydrophilic SiO₂ attracts the surfactant in certain areas. The weak attraction interaction forces also occur in the hydrophobic site of silica oxide by having a more hydrophobic (neutral) area around the SiO₂; the SiO₂ will attract other hydrophobic species, including the surfactant tail, through hydrophobic-hydrophobic interactions.

Figure 6c depicts the screened charge density of the SDBS surfactant. Two peaks in the hydrophobic and HB-acceptor region are typically found in the surfactant molecule. The alkyl chain or tail is depicted as a peak in the hydrophobicity region. The surfactant head group, Sulfonate (SO₃)⁻, comprised of negatively charged oxygen atoms, is depicted as a peak in the HB-acceptor region. Oxygen atoms in the headgroup contribute to the area of negative charge of the molecule surface.

Further, as all electron-rich oxygen atoms are concentrated in the head section of the SDBS, the negatively charged head section will orient the molecule towards the tiny area of positively charged hydrogen atoms in the silanol functional of SiO₂. At the same time, the presence

of negatively charged oxygen atoms on the silanols will act as barriers to repulse the negatively charged surfactant head group. Therefore, the probability of surfactants being adsorbed will depend on the presence of silanol function and the types of surfactant head groups.

The benzene ring between the head and tail section acts as a buffer separating the hydrophilic head from the hydrophobic carbon chain tail. Although the total charge of the benzene ring is zero, the sigma profile is extended from $-0.0075 \text{ e}/\text{\AA}^2$ to $0.0075 \text{ e}/\text{\AA}^2$. The benzene ring has rich inner ring electrons from freely moving electrons in π molecular orbital [72]. To balance the total charge of the molecule, the outer ring, which contains rich hydrogen atoms, consists of a more positive charge density. This unique charge density distribution will attract two benzene rings together in several probable configurations, including parallel, sandwich, T-shape, and Y-shape [73], increasing the probability for surfactant molecules to attract one another. Lastly, the hydrophobic tails interact with the hydrophobic section of the SiO_2 surface and the tail part from other surfactant molecules through weak hydrophobic-hydrophobic interactions. The longer the tail, the more hydrophobicity the surfactant behaves [74], as shown in Fig. 6c, which results in a higher probability of attracting the tail of the adjacent surfactants.

The information from screened charge density, molecule shape, and σ -profile was considered to interpret the possible molecular interactions between SiO_2 and surfactants. Several interactions are identified as follows:

- i. The repulsion force between negatively charged surfactant headgroup (SO_3^-) and valence oxygen (O^-) in the SiO_2 lattice.
- ii. The surfactant headgroup containing a negatively charged oxygen atom is attracted to the positively charged hydrogen atom on the silanols function of SiO_2 via hydrogen donor-acceptor interaction. This interaction is potentially limited due to the stronger repulsion of the interaction in (i).
- iii. The interaction between the benzene ring via π - π interaction between surfactant molecules potentially occurs in various configurations, resulting in the surfactant molecule's aggregation.
- iv. Hydrophobic attraction from the tail-tail interaction between surfactant hydrophobic tails enhances the probability of multilayer adsorption.

Model evaluation using temperature variation for isotherm prediction.

The converged models were tested by varying the concentrations and testing conditions to evaluate the consistency of the model and compare it to the reference

physical explanation. As the machine learning model works best in the interpolation problem, the temperature parameter was chosen as it has the highest range and variation in the tested dataset compared to other parameters, such as pH and salinity. To generate the adsorption prediction curve, SDBS from the surfactant dataset and sand with average properties of PZC of 7.08 and a specific area of $1.03 \text{ m}^2/\text{g}$ from the sand dataset were used [7, 40, 41, 45, 75]. The salinity of 0%wt and pH of 7 were selected, while the temperature was varied from 20°C to 50°C to be used as the input parameters.

In addition to the PINN model, the conventional neural network model was also simulated to be compared with the PINN model result. Three indicators were evaluated: (1) consistency: the adsorption result showed a consistent increase as the initial concentration increases; (2) saturation point: following the Langmuir isotherm, the plateau adsorption is attained, and additional surfactant concentrations will not increase the adsorption; (3) trend compared to the experimental results.

The results from both models are shown in Fig. 7. Although both neural network models have shown consistency, only PINN model achieved the saturation point. The conventional neural network model fails to comply at the highest concentration, and the saturation point is not yet achieved and tends to show consistent increases. Meanwhile, for the third indicator, adsorption typically decreases as temperature increases; as adsorption is a thermodynamic phenomenon, higher temperature leads to higher kinetic energy, and there is less probability of the molecule being attached to the solid surface [9]. A later study found that the adsorbed capacity depends on adsorption density; adsorption decreases for anionic surfactants with high adsorption density (entropy-driven) but contrasts for enthalpy-driven (low adsorption density like nonionic surfactants) [76]. The PINN model agrees with this third indicator for evaluating SDBS, an anionic surfactant. The PINN follows all indicators based on a mechanistic model. The Langmuir model, which was integrated into the neural network, was limiting the model's flexibility following the characteristic of the integrated mechanistic model.

Evaluation of PINN model with unseen surfactant

In this section, the model was challenged to predict the adsorption isotherm from two surfactants not used in the training or testing sets. Two surfactants were tested; the first surfactant is Internal Olefin Sulfonate (IOS), a sulfonate with double-branched anionic surfactant. This surfactant structure is completely new to the trained model. Meanwhile, the second surfactant is the Nonyl Phenol Ethoxylate-(NPE) series, a nonionic surfactant

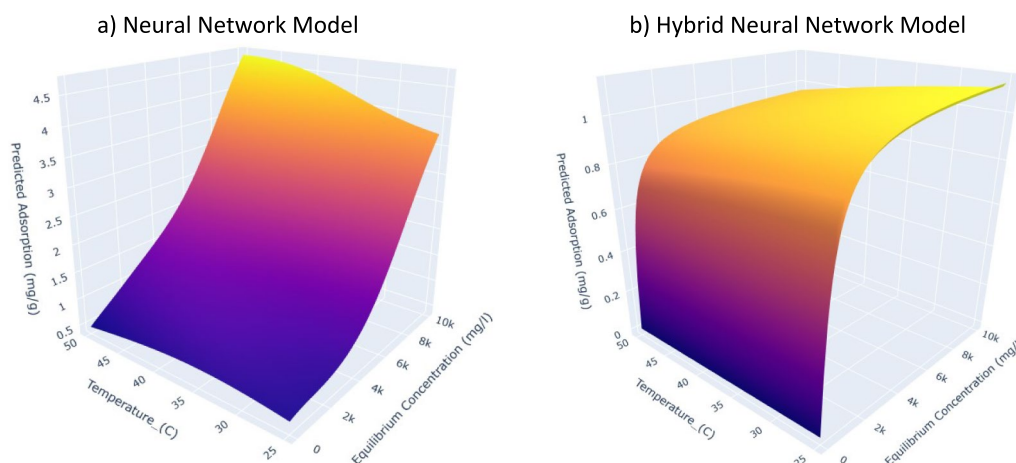


Fig. 7 **a** Surfactant adsorption isotherm generated from conventional neural network configuration with temperature variation **b** Surfactant adsorption isotherm generated from PINN model

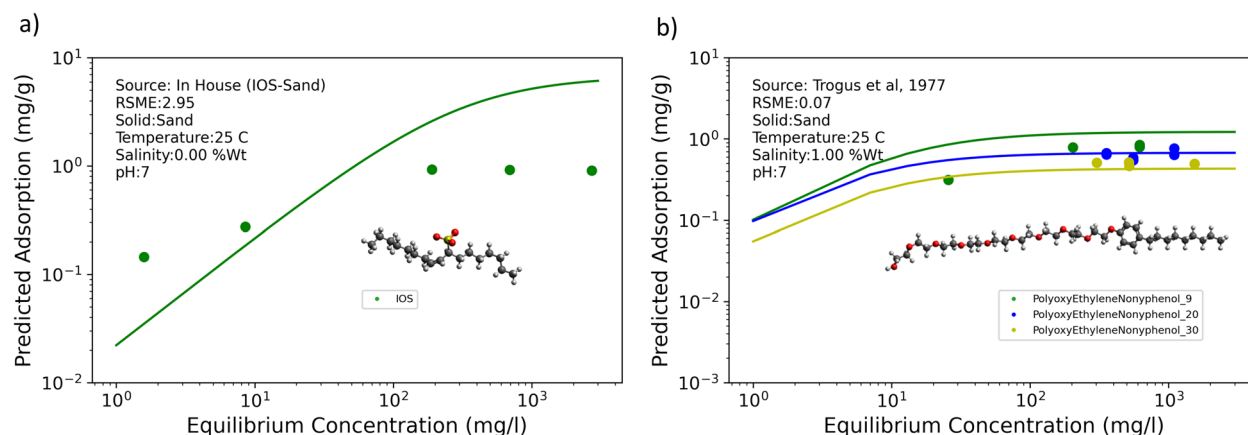


Fig. 8 Model evaluation using unseen surfactant **a** Internal Olefin Sulfonate (IOS) and **b** Polyoxyethylenononyphenol with 9, 20, and 30 EO groups. The dots represent the experimental adsorption values, and the line represents the model adsorption predicted value

with 9, 20, and 30 extended EO groups and a phenyl ring. The structure of this surfactant is found in non-ionic surfactants under the trade name of Triton-X and one type of NPE 20 already used in the training dataset. The adsorption result was compared to the experimental results from in-house laboratory testing and reference [77], using the testing condition and solid properties specified from the experimental data.

The result shown in Fig. 8 compares the adsorption prediction from the PINN model to the experimental result. Generally, the predicted isotherms onto the sand surface for IOS and NPE follow general characteristics from a Langmuir isotherm with adsorption capacity until the maximum amount of adsorbed surfactant is reached. The IOS adsorption prediction in Fig. 8a shows a higher adsorption capacity and maximum adsorption with an RMSE of 2.95 mg/g. This value is

relatively high compared to the NPE prediction adsorption in Fig. 8b, with an RMSE of only 0.07 mg/g. Even more, the prediction of nonionic surfactant indicates the correct trend of the maximum adsorption capacity with the increased number of EO groups following the experimental results. There are several factors affecting the differences in these results:

- i. The maximum adsorption of IOS surfactant is significantly higher than that of NPE, as shown in Fig. 8. The area per molecule of NPE is significantly higher than that of the IOS, compared with a similar adsorbent surface area, so fewer NPE molecules will fit on the surface. A higher adsorption magnitude for IOS also means the model has a higher absolute error of the prediction.

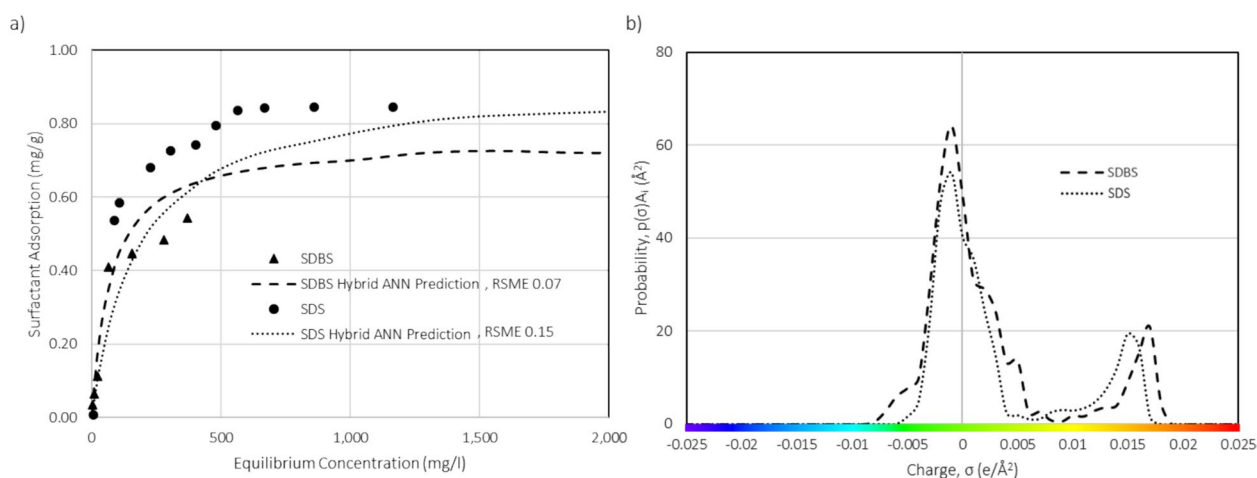


Fig. 9 **a** Sodium Dodecyl Benzene Sulfonate (SDBS) SO_3^- [9] and Sodium Dodecyl Sulfate (SDS) OSO_3^- [45] adsorption isotherm profile, **b** Screened charge density of SDBS and SDS

- ii. Compared to the nonionic surfactants, the anionic surfactant molecules vary widely in structures. For anionic surfactants, their structures are varied by head groups (e.g., carboxylate, sulfate, and sulfonate), extended function, presence of benzene ring, branch chain, and carbon tail length. Meanwhile, the ethoxylate-based nonionic surfactants only have an alcohol connecting with typically EO as a head group and a shorter carbon tail length (e.g., C12). The more molecular structure of the anionic compared to the nonionic surfactants, the more data variation there is for the idealized training set for the anionic surfactants.
- iii. The presence of NPE 20 in the training set provides an accurate starting point for the predicted adsorption model for slight changes in the molecular structure.
- iv. The IOS in the adsorption study was industrial grade with a relatively lower percentage of active ingredients than NPE. Also, the industrial-grade IOS generally contains average carbon numbers and internal branching. The effect of surfactant structural heterogeneity would result in lower IOS adsorption capacity than the predicted values due to more disordered packing of the adsorbed surfactants at the interface.

Surfactant structure analysis to adsorption behavior

This section further analyzes the surfactant adsorption dataset by examining the effect of the main structural component and sigma profile of the head group, tail length, and extended function.

Surfactant head group analysis

Surfactant adsorption isotherms of SDBS [9] and SDS [45] were examined to relate the screened charge density of the surfactant head group to the adsorption profile. SDBS and SDS were selected because their structure is the simplest compared to other types like branched or extended surfactant, and the alkyl length is similar, i.e., C12, so the effect other than the headgroup can be minimized. Also, both surfactants represent the main anionic surfactant class: sulfate (OSO_3^-) and sulfonate (SO_3^-).

Figure 9a illustrates the adsorption profile of both surfactants. SDS, the sulfate surfactant, has a higher adsorption capacity than SDBS, a sulfonate surfactant. The PINN model also showed an accurate trend with RMSE: 0.07–0.015 mg/g. The investigation was further extended by evaluating the screened charge density.

The screened charge density of SDS and SDBS are shown in Fig. 9b. There are two main peaks for both surfactants (1) the hydrophobicity region at charge density -0.005 to 0.005 $\text{e}/\text{\AA}^2$, which represents a neutral hydrocarbon tail section, and (2) the hydrogen bonding acceptor region at 0.012 to 0.017 $\text{e}/\text{\AA}^2$ which represents the rich electron oxygen atom on the headgroup. It should be noted that the sigma profile shows the positive number of screened charge density as it reflects the counter charge from a continuum perspective, where the actual molecule surface charge is negative. Although both screened charge density looks similar, the headgroup peak for SDBS is shifted more to the right side as it has a more intense charge density than SDS. This was due to less oxygen atoms in the sulfonate headgroup, which only contains three oxygen atoms compared to four in the sulfate. As the total charge for both surfactants is -1 , the headgroup, which contains less oxygen atoms, will have

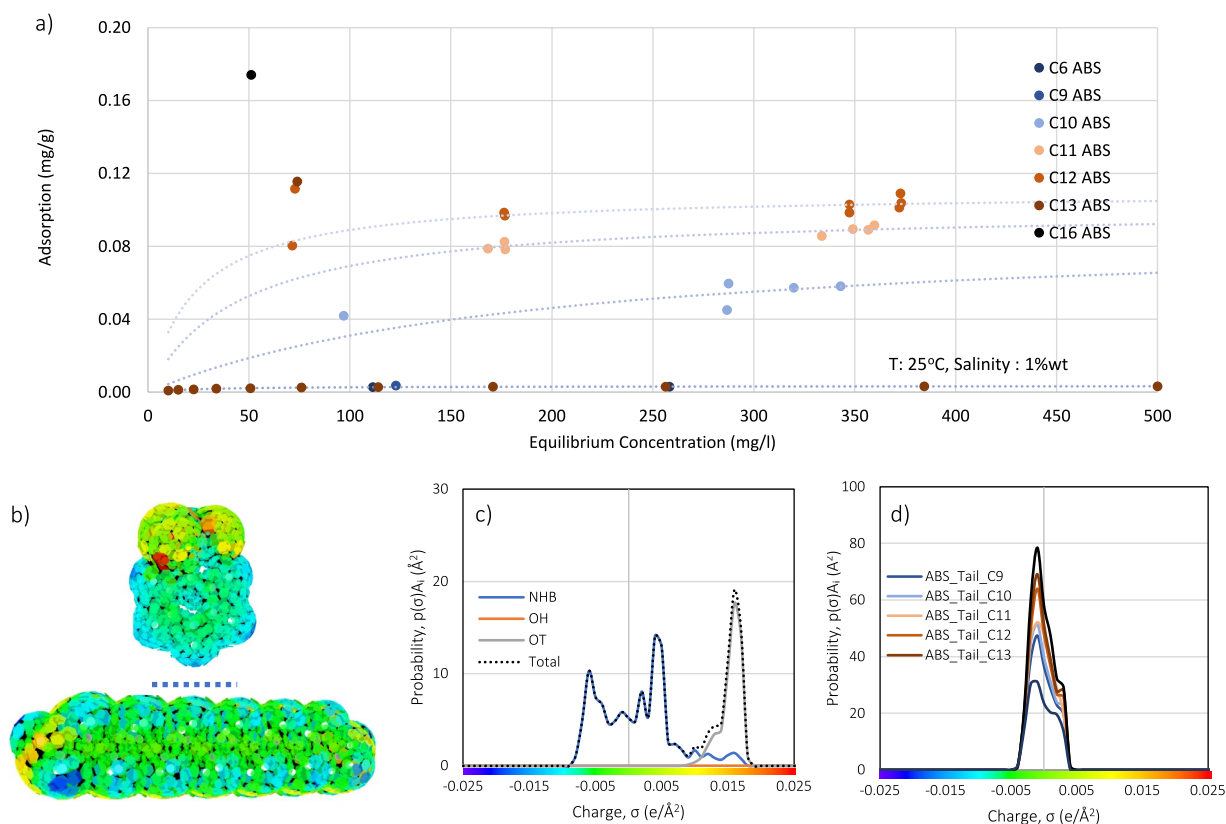


Fig. 10 **a** Adsorption isotherms for anionic surfactants, Alkyl Benzene Sulfonate (ABS) on Berea Sandstone: data points recreated from Trogus, Sophany [77] (symbols in legends) and predicted from this work (dotted lines), **b** Screened charge density of linear ABS (C-12) in section, **c** Sigma profile of linear ABS, and **d** Sigma profile various ABS tail lengths

a smaller negative region to distribute the charge, making the charge density of the sulfonate headgroup more intense.

In COSMO, the y-axis represents the probability of finding a specific charge, and the x-axis is the charge's density. As the charge is seen from a continuum perspective, the peak in the extreme positive charge density means the intense negative charge on the actual molecule surface [78]. Related to the adsorption isotherm, a more intense charge density in SDBS means the molecule has a denser negative surface. As the silica oxide surface is negatively charged, a higher intense negatively charged headgroup will have a more substantial probability of repulsing the solid surface, avoiding adsorption. Meanwhile, the SDS with a less dense negative surface will have a higher probability of being adsorbed on the silica oxide surface.

The hydrophobicity region of these two surfactants also interacts with the neutral region of silica oxide, resulting in hydrophobic-hydrophobic interactions. However, since both hydrophobicity peaks are comparable in size and location, there is no trend to evaluate. Another effect

on the hydrophobicity region will be examined, which is the effect of surfactant tail length.

Surfactant tail length

Trogus, Sophany [77] examined the adsorption of alkyl benzene sulfonate (ABS) as it is used as a similar type of surfactant with tail length variation. The media used was Berea Sandstone with 1%wt brine solution (NaCl). Figure 10a shows that the longer the alkyl chain of the surfactant molecule, the higher the adsorption it reveals. The study also mentioned that the surfactant adsorption with C6 and C9 alkyl chain lengths are relatively low as they have better water solubility than the longer alkyl chain lengths.

In Fig. 10b, ABS has a benzenesulfonate headgroup followed by a linear hydrocarbon tail. The valence electron charge (-1) of oxygen was distributed along the sulfonate headgroup surface, making the surface negative. The surface is positive in the sigma profile as viewed from the COSMO medium perspective and is colored orange.

Furthermore, this surface is interpreted as the OT hydrogen bonding and acts as the hydrogen bonding

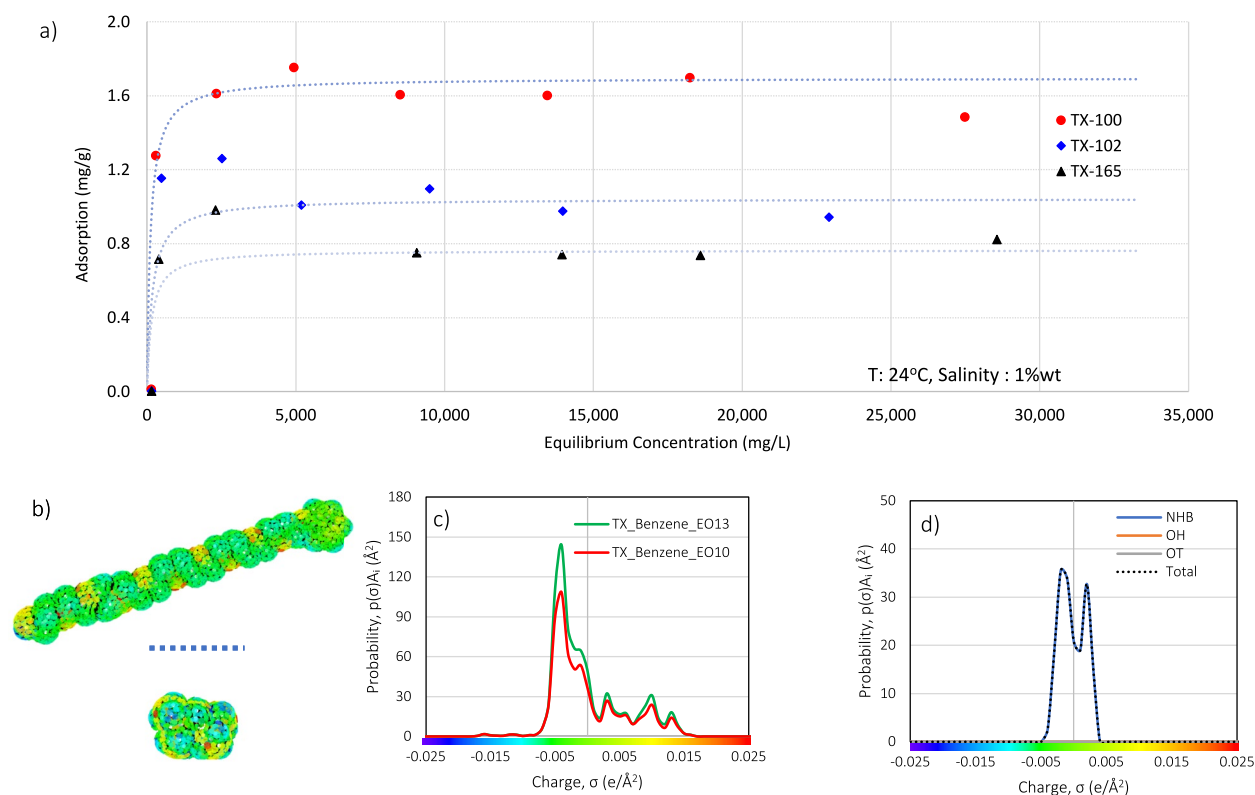


Fig. 11 **a** The adsorption of Triton-X series surfactant onto crushed Berea Sandstone: data points recreated from Lawson [43] (symbols in legends) and predicted from this work (dotted lines), **b** the screened charge density of Triton-X 100 in section, and **c** sigma profile of the heads, and **d** tails of Triton sulfonate with EO-10 and EO-13

acceptor. In the sigma profile shown in Fig. 10c. the sulfonate is represented by a peak at 0.015 (e/Å²). The neutral region represents the benzene function from the headgroup with a peak around 0.005 and -0.005 (e/Å²). Since the headgroup from all surfactants is the same, there is no effect on adsorption to evaluate. As shown in Fig. 10d, the longer the tail length, the higher the peak in the hydrophobicity region is observed. The hydrophobicity region generally represents the neutral region that does not like water [62]. The trend in the sigma profile agrees with surfactant solubility to water. The higher the hydrophobicity peak, the less soluble surfactant dissolves in water, making the tendency of surfactant molecules to aggregate from tail–tail interactions. The hydrophobicity region is also in line with the misfit energy interactions. This interaction represents weak interaction from different weak charges around the molecule surface. The higher the hydrophobicity peaks, the more potential the molecules form hydrophobic-hydrophobic interactions with the hydrophobic region on the solid, making the adsorption greater.

Surfactant extended function

The adsorption profiles of NPE or Triton-X series (TX), reported by Lawson [43], with EO ranging from 9.5 to 30, were evaluated. High-purity surfactant was used in their work. The media was sandstone, and 1 wt.% Sodium Chloride solution was used. The static experiment results are shown in Fig. 11a. The conclusion of this study showed that the adsorption onto sandstone depends on the extent of the EO number. Also, the adsorption of TX is relatively insensitive to solution salinity. Furthermore, the solubility of EO increases with the increase in EO number.

The screened charge density for TX is shown in Fig. 11b. TX has an alcohol head group followed by an EO extended function. Because of its weakly charged non-ionic head group, TX has an overall greenish color, showing that the charge disparity for the whole section is not in great contrast. As the total charge is zero, the surface charge has been determined by the presence of an atom with higher electron valence, such as oxygen. Excess electrons on oxygen will provide a HB-acceptor region. This region has a peak of around 0.010 to 0.015 (e/Å²), as seen in Fig. 11c. The hydroxyl function in alcohol also shapes

the hydrogen bonding donor peaks but is relatively low at around -0.015 (e/A^2).

The EO group of TX surfactant contributed to the hydrophobicity region and HB-acceptor region. Figure 11c shows the differences between TX's EO10 and EO13 sigma profiles. An additional number of EO contributes to two atoms of carbon, which increases the peak in hydrophilicity, whereas an oxygen atom increases the peak in the HB-acceptor region. Out of five series, only EO10 and 13 were simulated in this study. The rest cannot be simulated due to the maximum number of atoms run by GAMESS.

Compared to the findings in the reference study, the higher surfactant solubility with higher EO can be described by more HB-acceptor region on the molecule surface. However, the study by Di Serio and Tesser [79] showed that for surfactants with very high EO numbers, the EO group reduced the molecule solubility to a higher EO number. The peak ratio between hydrophobicity and HB-acceptor for each additional EO group can describe this phenomenon. It is suggested that there is a maximum ratio of hydrophobicity and HB-acceptor where the solubility is immensely reduced.

On the adsorption side for surfactants with very high EO numbers, the study concluded that the lower the number of EO groups, the higher its adsorption. Besides the effect of surfactant solubility, higher adsorption due to EO can be described using the COSMO concept. The more the EO group in the surfactant molecule, the more HB-donor contribution is obtained, increasing the interaction between the surface and surfactant molecules. As the HB-acceptor region distributes to the whole length of the extended function of the surfactant, the interaction could affect the orientation of the adsorbed surfactant molecules onto the solid surface. This orientation will cover more HB-donor regions on a solid surface, which reduces the tendency for further adsorption. However, the orientation information is not captured in the sigma profile, and a similar orientation also occurs for the surfactants with a higher number of EO.

The reduction of adsorption of the surfactant with higher EO can be described from the hydrophobicity peak, which shifted to the HB-donor region. The TX hydrophobicity peak is concentrated on negative screened charge density, which means that the molecule has an excess donor region; the probability of this excess donor region is higher with an increasing number of EO groups. Silica oxide has an excess hydrogen donor on the counterpart, resulting in incompatible misfit energy interactions. Misfit is the electrostatic interaction between two molecules arising from the imperfect fit of two polarities [78]. Though it is weak compared to hydrogen bonding, the peak is considerably high.

The tail part of the TX series consists of a similar-length hydrocarbon chain. This segment then accumulated in the hydrophobicity range of the sigma profile curve. As shown, there is no variation in the tail length of the TX series; there is no further analysis describing the effect of tail length.

SHAP evaluation on model feature

The SHAP (SHapley Additive exPlanations) analysis is utilized to illustrate the importance and impact of individual features on the model's predictions. Figure 12a shows how each feature contributes to the model output by shifting the predicted value from the base prediction, $E[f(x)] = 1.06$ (the mean prediction across the sample set), to a final model prediction of 0.84, while Fig. 12b shows how the SHAP value attributes to the feature globally. According to the SHAP evaluation, the trained model considers the most influential features determining adsorption are: (1) molecular structure, represented by molecular volume (MolVol) and length; (2) screened charge density of both the head and tail groups; (3) equilibrium concentration; and (4) solution conditions.

The molecular cavity volume or MolVol is the most important feature with the largest negative mean SHAP value. This descriptor is the intermediate output of GAMESS/COSab when calculating COSMO screened charge density. This value refers to molecular cavity in the solvation continuum when molecular surface is constructed. Larger molecular cavity volumes lead to lower predicted adsorption. The larger molecules tend to have limited interaction or accessibility due to steric hindrance, especially those with bulkier groups or branched chains, resulting in higher occupation on the surface of the adsorbent. This steric effect can hinder the formation of closely packed surfactant molecules, especially in systems where head group repulsion is significant.

Tail hydrophobicity marked as Tail_NHB-Hydrophobicity in Fig. 12 also enhances adsorption, which correlates with the tail length, since longer hydrocarbon chains contribute more to overall hydrophobicity. This is due to the stronger van der Waals interactions between the hydrocarbon chains and the surface or nearby adsorbed surfactant molecules. This finding supports the discussion in Sect. "Surfactant tail length" on surfactant tail length. Additionally, the presence of hydrogen bond acceptors (NHB-HB_Acceptor) in the head group negatively affects adsorption. This aligns with observations in Sect. "Surfactant extended function" on extended surfactant functions; an increase in EO (ethylene oxide) groups enhances solubility in water, making adsorption onto the surface less likely. When the NHB-HB_Acceptor value of the head group increases significantly, such as in surfactants with a high number of ethylene oxide (EO)

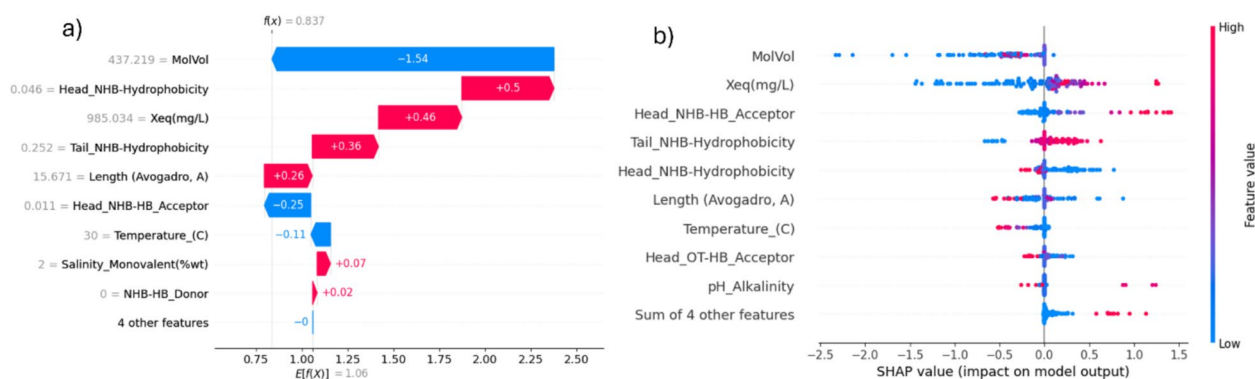


Fig. 12 **a** Waterfall plot of SHAP values for an individual prediction from the PINN model. Features in blue decrease the prediction (push $f(x)$ lower), while features in red increase it (push $f(x)$ higher). **b** SHAP summary plot showing the distribution of SHAP values for each feature across all data points. The color represents the feature value (red = high, blue = low), and the position on the X-axis shows the direction and magnitude of the feature's impact on the adsorption capacity. Note: Xeq is equilibrium concentration, in mg/L; MolVol is total volume cavity of molecule from GAMESS/COSab result in $\text{\AA}^3$; and the physical meaning of COSMO feature is described in Sect. "Physical meaning of selected descriptors"

units, the SHAP values reflect a positive contribution to adsorption. This trend is likely due to increased hydrophobic character and reduced water solubility associated with higher EO content. This result aligns well with the case studies in Sect. "Surfactant extended function".

Solution conditions, such as temperature and salinity, also influence adsorption behavior. As discussed in Sect. "Model evaluation using temperature variation for isotherm prediction.", higher temperatures typically reduce adsorption due to increased molecular kinetic energy, while higher salinity can enhance adsorption through precipitation effects. Finally, tail length and hydrophobicity parameters both show strong influence, underscoring the importance of molecular size and surface interactions in the adsorption process.

Conclusion

This study developed a deep learning model to predict anionic and nonionic surfactant adsorption onto silica surfaces. The model covers several surfactant head groups, including sulfate, sulfonate, sulfosuccinate, and alcohol, with various alkyl and extended groups. Several molecular descriptors were generated and tested in the RFE to be selected as input parameters for the deep learning model. The surfactant descriptors using COSMO-screened charge density resulted in the highest accuracy among several surfactant descriptors. This work demonstrated that the COSMO-screened charge density and σ -profile can be used to reveal probable interactions between the surfactant and the silica surface, and the interpretation relates well with the experimental result. The incorporation of the Langmuir adsorption isotherm into the deep learning model in our work, so-called the PINN model, provided a more logical rationale and

higher accuracy results than the solely neural network model as it embeds the classical Langmuir characteristics with adsorption capacity and maximum adsorption for the surfactant systems. The performance of our PINN model on the prediction of unseen surfactant gives reasonably good results in the surfactant with common molecular structure as appeared in the training dataset; however, the prediction still needs improvement in the case of the completely new structure with the heterogeneity of the molecular structure. Further, the SHAP analysis also show consistent results from study cases to evaluate descriptors impact to the adsorption. This work can provide guidance on screening surfactant formulations to minimize the adsorption capacity before applying the formulations in the real field in EOR or remediation projects.

Future work should focus on extending the surfactant database, molecular representation and machine learning model. The adsorption isotherm database shall be extended to various conditions, especially in extreme regions such as high salinity or high temperature, where this is the typical condition in the oil reservoir. A validity check can also be done at this high salinity high temperature condition. In terms of molecular representation, the screened charge density from the whole molecule surfactant will give a more accurate screened charge density and sigma profile value. Several studies have shown the potential of the generative models (such as variational autoencoder, VAE) in predicting larger molecular properties, like protein and RNA. Therefore, in the end, we envision the combination of the complete database and strong molecular representation shall improve the ability of our model to predict surfactant adsorption and improve the model's performance and efficiency.

Abbreviations

ABS	Alkyl Benzene Sulfonate
ANN	Artificial neural networks
AOT	Bis(2-ethylhexyl)sulfosuccinate
CMC	Critical micelle concentration
COSMO	Conductor-like Screening Model
DHSS	Bis(2-ethylhexyl)sulfosuccinate
DOWFAX	Alkydiphenyloxide Disulfonate
EACN	Equivalent Alkane Carbon Number
EO	Ethylene Oxide
EOR	Enhanced oil recovery
GC	Group contribution
HB	Hydrogen bonding
IFT	Interfacial Tension
IOS	Internal Olefin Sulfonate
MolVol	Total volume cavity of molecule from GAMESS/COSab result, in Å ³
Neodol	A surfactant trade name of Alkyl Alcohol Ethoxylate
NHB	Non-hydrogen bonding
NPE	Nonyl Phenol Ethoxylate
OSO ₃ ⁻	Sulfate head group
OT	Other
PINN	Physics-Informed Neural Network
PO	Propylene Oxide
PZC	Point of Zero Charge
RFE	Recursive feature elimination
RHF	Restricted Hartree–Fock
RMSE	Root means square error
SDBS	Sodium Dodecyl Benzene Sulfonate
SDS	Sodium Dodecyl Sulfate
SLES	Sodium Lauryl Ethyl Sulfate
SMILES	Simplified molecular-input line-entry system
SO ₃ ⁻	Sulfonate head group
Tergitol	A surfactant trade name for Alkylphenol Ethoxylate
Triton-X	A surfactant trade name for Nonyl Phenol Ethoxylate
TX	Triton-X
Xeq	Equilibrium concentration of surfactant adsorption, in mg/l

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13321-025-01027-y>.

Supplementary material 1

Acknowledgements

The authors would like to express gratitude to The Petroleum and Petrochemical College, Center of Excellence on Petrochemical and Materials Technology (PETROMAT), Chulalongkorn University, and the graduate scholarship program for ASEAN or Non-ASEAN countries, Chulalongkorn University for scholarship support. The work was financially supported by the Thailand Science Research and Innovation Fund Chulalongkorn University (Grant No. DIS66630029).

Author contributions

A.A.A. wrote the main manuscript text and worked on modeling, K.S. and W.Y. worked with data curation and validation, A.C. and B.J.S. conceptualized and reviewed the manuscript, US conceptualized, reviewed, and supervised the daily work.

Funding

This research was funded by the Thailand Science Research and Innovation Fund Chulalongkorn University (DIS66630029).

Data availability

No datasets were generated or analysed during the current study.

Declarations**Ethics approval and consent to participate**

This declaration is not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹The Petroleum and Petrochemical College, Chulalongkorn University, Bangkok, Thailand. ²Center of Excellence On Petrochemical and Materials Technology (PETROMAT), Chulalongkorn University, Bangkok, Thailand. ³Mewbourne School of Petroleum and Geological Engineering, University of Oklahoma, Norman, OK, USA.

Received: 6 August 2024 Accepted: 6 May 2025

Published online: 26 May 2025

References

- Rosen MJ, Kunjappu JT (2012) Surfactants and interfacial phenomena. Wiley, Hoboken
- Massarweh O, Abushaikh AS (2020) The use of surfactants in enhanced oil recovery: a review of recent advances. *Energy Rep* 6:3150–3178
- Mao X et al (2015) Use of surfactants for the remediation of contaminated soils: a review. *J Hazard Mater* 285:419–435
- Liu Q et al (2004) Improved oil recovery by adsorption–desorption in chemical flooding. *J Petrol Sci Eng* 43:75
- Amirianshoja T et al (2013) A comparative study of surfactant adsorption by clay minerals. *J Petrol Sci Eng* 101:21–27
- Belhaj AF et al (2020) The effect of surfactant concentration, salinity, temperature, and pH on surfactant adsorption for chemical enhanced oil recovery: a review. *J Petrol Explor Prod Technol* 10(1):125–137
- Mushtaq M et al (2015) Effect of pH on the static adsorption of foaming surfactants on Malaysian sandstone. *Arab J Geosci* 8(10):8539–8548
- Karin Mannhardt, L.L. Schramm, and J.J. Novosad, *Effect of rock type and brine composition on adsorption of two foam-forming surfactants*. SPE Adv Technol Ser, 1993. 1: 212
- Ziegler VM, Handy LL (1981) Effect of temperature on surfactant adsorption in porous media. *Soc Petrol Eng J* 21(02):218–228
- Kalam S et al (2021) Surfactant adsorption isotherms: a review. *ACS Omega* 6(48):32342–32348
- Somasundaran P, Middleton R, Viswanathan K (1984) Relationship between surfactant structure and adsorption. *Struct Perform Relation Surf* 253:269
- Atkin R et al (2003) Mechanism of cationic surfactant adsorption at the solid–aqueous interface. *Adv Coll Interface Sci* 103(3):219–304
- Marrero J, Gani R, Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 2001(183–184).
- Talu O, Zwiebel I (1986) Group-contribution based estimation of pure component properties. *AIChE J* 32:1263
- Constantinou L, Gani R (1994) New group contribution method for estimating properties of pure compounds. *AIChE J* 40(10):1697–1710
- Mattei M, Kontogeorgis GM, Gani R (2013) Modeling of the critical micelle concentration (CMC) of nonionic surfactants with an extended group-contribution method. *Ind Eng Chem Res* 52(34):12236–12246
- Guo X, Rong Z, Ying X (2006) Calculation of hydrophile–lipophile balance for polyethoxylated surfactants by group contribution method. *J Colloid Interface Sci* 298(1):441–450
- Wu J et al (2004) Development of a method for measurement of relative solubility of nonionic surfactants. *Colloids Surf A* 232(2–3):229–237
- Mattei M, Kontogeorgis GM, Gani R (2014) A comprehensive framework for surfactant selection and design for emulsion based chemical product design. *Fluid Phase Equilib* 362:288–299
- Chanachichalermsong W, Charoensang A, Suriyapraphadilok U (2019) Krafft point prediction of anionic surfactants using group contribution method: first-order and higher-order groups. *J Surfactants Deterg* 22(4):907–919
- Huibers PDT (1999) Quantum-chemical calculations of the charge distribution in ionic surfactants. *Langmuir* 15(22):7546–7550
- Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J Phys Chem* 99(7):2224–2235

23. Lukowicz T et al (2015) Rationalization and prediction of the equivalent alkane carbon number (EACN) of polar hydrocarbon oils with COSMO-RS σ -moments. *Langmuir* 31(41):11220–11226
24. Amar Y et al (2019) Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. *Chem Sci* 10(27):6697–6706
25. Rittig JG, et al. Graph neural networks for temperature-dependent activity coefficient prediction of solutes in ionic liquids. *arXiv preprint arXiv:2206.11776*, 2022.
26. Soares RDP (2011) The combinatorial term for COSMO-based activity coefficient models. *Ind Eng Chem Res* 50(5):3060–3063
27. Hsieh C-M, Sandler SI, Lin S-T (2010) Improvements of COSMO-SAC for vapor–liquid and liquid–liquid equilibrium predictions. *Fluid Phase Equilib* 297(1):90–97
28. Hsieh C-M, Lin S-T, Vrabec J (2014) Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior. *Fluid Phase Equilib* 367:109–116
29. Sellaoui L et al (2016) Binary adsorption isotherms of two ionic liquids and ibuprofen on an activated carbon cloth: simulation and interpretations using statistical and COSMO-RS models. *RSC Adv* 6(72):67701–67714
30. Sellaoui L et al (2017) New insights into the adsorption of crystal violet dye on functionalized multi-walled carbon nanotubes: experiments, statistical physics and COSMO-RS models application. *J Mol Liq* 248:890–897
31. Sellaoui L et al (2016) Steric and energetic interpretations of the equilibrium adsorption of two new pyridinium ionic liquids and ibuprofen on a microporous activated carbon cloth: statistical and COSMO-RS models. *Fluid Phase Equilib* 414:156–163
32. Pelalak R et al (2021) Molecular dynamics simulation of novel diamino-functionalized hollow mesosilica spheres for adsorption of dyes from synthetic wastewater. *J Mol Liquids* 322:114812
33. Toumi K-H et al (2018) Molecular modeling of cationic dyes adsorption on agricultural algerian olive cake waste. *J Mol Liq* 264:127–133
34. Klamt A, Eckert F, Diedenhofen M (2002) Prediction of soil sorption coefficient with a conductor-like screening model for real solvents. *Environ Toxicol Chem* 12(12):2562–2566
35. Limousin G et al (2007) Sorption isotherms: a review on physical bases, modeling and measurement. *Appl Geochem* 22(2):249–275
36. Allen SJ, McKay G, Porter JF (2004) Adsorption isotherm models for basic dye adsorption by peat in single and binary component systems. *J Colloid Interface Sci* 280(2):322–333
37. Malek A, Farooq S (1996) Comparison of isotherm models for hydrocarbon adsorption on activated carbon. *AIChE J* 42(11):3191–3201
38. Kalam S, et al, Soft computing approach for the prediction of surfactant adsorption, in ADIPEC. 2022. p. D022S155R002.
39. Chen F et al (2021) On the evaluation of rhamnolipid biosurfactant adsorption performance on amberlite XAD-2 using machine learning techniques. *Biomed Res Int* 2021:5530093
40. Bera A et al (2013) Adsorption of surfactants on sand surface in enhanced oil recovery: Isotherms, kinetics and thermodynamic studies. *Appl Surf Sci* 284:87–99
41. Azam MR et al (2013) Static adsorption of anionic surfactant onto crushed Berea sandstone. *J Petrol Explor Prod Technol* 3(3):195–201
42. Kwok W et al (1993) Static and dynamic adsorption of a nonionic surfactant on Berea sandstone. *Colloid Surf* 78:193
43. Lawson JB, The adsorption of nonionic surfactant on sandstone and carbonate. *SPE J*, 1978(7052).
44. Ngo I et al (2019) Effects of reversibility on enhanced oil recovery using sodium dodecylbenzene sulfonate (SDBS). *J Jpn Petrol Inst* 62(4):188–198
45. Gong W et al (2015) Properties of surfactants on high salt-affected sandy land in enhanced sand fixation: salt tolerance, adsorption isotherms and ecological effect. *RSC Adv* 5(100):81847–81856
46. C, Effect of silica nanoparticle on surfactant adsorption onto silica and alumina oxide surfaces for enhanced oil recovery applications, in the petroleum and petrochemical college. 2020, Chulalongkorn University: Thailand.
47. Miñana-Perez M et al (1995) Solubilization of polar oils with extended surfactants. *Colloids Surf A* 100:217–224
48. Postmus BR, Leermakers FAM, Stuart MAC (2008) Self-consistent field modeling of nonionic surfactants at the silica–water interface: incorporating molecular detail. *Langmuir* 24(8):3960–3969
49. Arpornpong N et al (2010) Ethoxy carboxylate extended surfactant: micellar, adsorption and adsolubilization properties. *J Surfactants Deterg* 13(3):305–311
50. Rattanaudom P et al (2023) Experimental investigation of hydrophobic and hydrophilic silica nanoparticles on extended surfactant properties: Micro-emulsion, viscosity, and adsorption behaviors. *Geoenerg Sci Eng* 223:211582
51. Hocine S, et al. Long term thermal stability of chemical EOR surfactants. 2018. SPE.
52. Nagarajan R (2002) Molecular packing parameter and surfactant self-assembly: the neglected role of the surfactant tail. *Langmuir* 18(1):31–38
53. El-Nahhal Y, Safi J (2005) Adsorption of benzene and naphthalene to modified montmorillonite. *J Food Agric Environ* 3(2):295–298
54. Shen Y-H (2002) Removal of phenol from water by adsorption–flocculation using organobentonite. *Water Res* 36(5):1107–1114
55. Stevens JJ, Anderson SJ (1996) Orientation of trimethylphenylammonium (TMPA) on Wyoming montmorillonite: implications for sorption of aromatic compounds. *Clays Clay Miner* 44:132–141
56. Man MS et al (2017) Screening cation and anion of ionic liquid for dissolution of silicon dioxide using COSMO-RS. *Ind J Sci Technol* 10:1–6
57. Gasteiger J, Marsili M (1979) Iterative partial equalization of orbital electronegativity-a to atomic charges. *Tetrahedron* 36:3219
58. Geoffrey H. Avogadro partial charges. 2012; Q&A Section for Avogadro. Available from: <https://discuss.avogadro.cc/t/partial-charges/248>. Accessed 14 Mar 2022.
59. Barca GMJ et al (2020) Recent developments in the general atomic and molecular electronic structure system. *J Chem Phys* 152(15):154102
60. Hanwell MD et al (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminf* 4(1):17
61. Bell IH et al (2020) A benchmark open-source implementation of COSMO-SAC. *J Chem Theory Comput* 16(4):2635–2646
62. Klamt A, Eckert F, Arlt W (2010) COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annu Rev Chem Biomol Eng* 1:101–122
63. Pedregosa F, et al., Scikit-learn: Machine learning in python. the Journal of machine learning research, 2011. **12**: p. 2825–2830.
64. Chen X, Jeong JC. Enhanced recursive feature elimination. in Sixth International Conference on Machine Learning and Applications (ICMLA 2007). 2007.
65. Abadi, M., et al., *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*. *arXiv preprint arXiv:1603.04467*, 2016.
66. Chollet F. Keras: the python deep learning library. *Astrophysics source code library*, 2018: p. ascl-1806.
67. Wesson LL, Harwell JH (2000) Surfactant adsorption in porous media. In: Schramm LL (ed) *Surfactants: fundamentals and application in petroleum industry*. Cambridge University Press, Cambridge
68. Zhuravlev LT (2000) The surface chemistry of amorphous silica. *Zhuravlev model Colloid Surf* 173:1–28
69. Kosmulski M (2002) The pH-dependent surface charging and the points of zero charge. *J Colloid Interface Sci* 253(1):77–87
70. Sulpizi M, Gaigeot M-P, Sprik M (2012) The silica-water interface: how the silanols determine the surface acidity and modulate the water properties. *J Chem Theory Comput* 8(3):1037–1047
71. Das SK, Bhattacharyya AJ (2010) Influence of oxide particle network morphology on ion solvation and transport in "soggy sand" electrolytes. *J Phys Chem B* 114(20):6830–6835
72. Moskowitz JW, Barnett MP (2004) Pi electronic structure of the benzene molecule. *J Chem Phys* 39(6):1557–1560
73. Headen TF et al (2010) Structure of π – π interactions in aromatic liquids. *J Am Chem Soc* 132(16):5735–5742
74. Hanson MG et al (2020) Effects of hydrophobic tail length variation on surfactant-mediated protein stabilization. *Mol Pharm* 17(11):4302–4311
75. Paternina CA et al (2020) Influence of salinity and hardness on the static adsorption of an extended surfactant for an oil recovery purpose. *J Petrol Sci Eng*. <https://doi.org/10.1016/j.petrol.2020.107592>
76. Kamal MS, Hussein IA, Sultan AS (2017) Review on surfactant flooding: phase behavior, retention, IFT, and field applications. *Energy Fuels* 31(8):7701–7720

77. Trogus FJ et al (1977) Static and dynamic adsorption of anionic and non-ionic surfactants. *SPE J.* <https://doi.org/10.2118/6004-PA>
78. Klamt A (2018) The COSMO and COSMO-RS solvation models. *WIREs Comput Mol Sci* 8(1):e1338
79. Di Serio M et al (1995) Ethylene oxide solubility and ethoxylation kinetics in the synthesis of nonionic surfactants. *Ind Eng Chem Res* 34(11):4092–4098

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.